TEXAS A&M UNIVERSITY

COLLEGE STATION, TEXAS 77843-3143

A TOEPLITZ GRAM-SCHMIDT ALGORITHM

FOR AUTOREGRESSIVE MODELING

Herbert T. Davis

Data Information Associates

Albuquerque

H. Joseph Newton

Institute of Statistics

Texas A&M University

Marcello Pagano

Department of Biostatistics

Harvard University

DTIC
SELECTED
JUL 1 9 1982

F

AD A117071

DTIC FILE COPY

82 07 19 008

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N-33	2. GOVT ACCESSION NO. AD-A117071	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A TOEPLITZ GRAM-SCHMIDT ALGORITHM FOR AUTOREGRESSIVE MODELING		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Herbert T. Davis, H. Joseph Newton, and Marcello Pagano		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-82-MP-2001 ARU DAAG 29080-C-0070
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Institute of Statistics College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE July 1982
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Yule-Walker Estimators, Modified Gram-Schmidt Decomposition, Levinson's Algorithm, Robust Autoregression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An algorithm is presented for efficiently finding the information needed in the modified Gram-Schmidt decomposition of the augmented autoregressive design matrix to find the Yule-Walker estimators of autoregressive parameters for orders 1, 2,...,M. The algorithm is shown to be slightly slower than Levinson's algorithm and to require slightly more storage but enjoys the superior numerical properties of the modified Gram-Schmidt decomposition algorithm. Further, it is shown how the algorithm provides a unified framework for suggesting robust autoregressive estimators, for finding autoregression diagnostics, and for understanding the Burg algorithm.

1. INTRODUCTION

Let $\{y(t), t\epsilon Z\}$ be a zero mean, covariance stationary, discrete time
series with autocovariance function $\{R(v), v\epsilon Z\}$. Then in: 1) finite order
autoregressive modeling (Akaike (1969)), 2) spectral density estimation
using autoregressive approximants (Parzen (1974)), 3) maximum entropy
spectral analysis (Ulrych and Bishop (1974)), and 4) explicit minimum mean
squared error linear prediction (see Anderson (1971), p. 419), one is faced
with estimating the quantities $\alpha_k(j), \sigma_k^2$, $1 \leq j \leq k \leq M$, satisfying the Yule-Walker
equations

$$\sum_{j=0}^{k} \alpha_k(j) R(j-v) = \delta_v \sigma_k^2 \quad , \quad v \geq 0 \quad , \tag{1.1}$$

where $\alpha_k(0) = 1$ and δ_v is the Kronecker delta.

Let $y(1),\ldots,y(T)$ be a mean detrended sample realization from y. The
traditional method for finding estimators $\hat{\alpha}_k(j)$ and $\hat{\sigma}_k^2$ of $\alpha_k(j)$ and σ_k^2 has
been to estimate $R(0),\ldots,R(M)$ by the positive definite sample autocovariances

$$\hat{R}(v) = \frac{1}{T} \sum_{t=1}^{T-|v|} y(t)y(t+|v|) \quad , \quad |v| \leq M,$$

and then to use (1.1) with $\hat{R}(v)$ replacing $R(v)$. These Yule-Walker estimators
have two main attractions; 1) they can be calculated very quickly by
Levinson's (1947) algorithm

$$\hat{\alpha}_1(1) = -\hat{R}(1)/\hat{R}(0) \quad , \quad \hat{\sigma}_1^2 = \hat{R}(0)[1-\hat{\alpha}_1^2(1)], \tag{1.2}$$

$$\hat{a}_k(k) = -[\sum_{j=0}^{k-1} \hat{a}_{k-1}(j)\hat{R}(k-j)]/\hat{\sigma}^2_{k-1} \quad ,$$

$$\hat{a}_k(j) = \hat{a}_{k-1}(j) + \hat{a}_k(k)\hat{a}_{k-1}(k-j), \qquad j=1,\ldots,k-1, \qquad (1.3)$$

$$\hat{\sigma}^2_k = \hat{\sigma}^2_{k-1} [1-\hat{a}^2_k(k)] \quad ,$$

$k=2,\ldots,M$, and 2) they are guaranteed to lead to stable prediction filters and positive spectral density estimators since (Parzen (1961)) the zeros of the complex polynomials $\hat{g}_k(z) = \sum_{j=0}^{k} \hat{a}_k(j)z^j$ are all outside the unit circle.

If we define the (k x k) Toeplitz matrix $\hat{\Gamma}_k$ to have (i,j)th element $\hat{R}(|i-j|)$ and the (k x 1) vector $\hat{\underline{r}}_k$ to have i^{th} element $\hat{R}(i)$, we can write the sample Yule-Walker equations for $v=1,\ldots,k$ as

$$\hat{\Gamma}_k \hat{\underline{a}}_k = -\hat{\underline{r}}_k \quad , \qquad (1.4)$$

where $\hat{\underline{a}}_k = (\hat{a}_k(1),\ldots,\hat{a}_k(k))^T$. Then (1.4) are the normal equations for the ordinary least squares estimators $\hat{\underline{a}}_k$ in the linear model

$\underline{y}_k = -X_k\underline{a}_k + \underline{\epsilon}$, where

$$\underline{y}_k = \begin{bmatrix} y(1) \\ y(2) \\ y(3) \\ \vdots \\ y(k+1) \\ \vdots \\ y(T) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad X_k = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ y(1) & 0 & \ldots & 0 \\ y(2) & y(1) & \ldots & 0 \\ \vdots & & & \\ y(k) & y(k-1) & \ldots & y(1) \\ \vdots & & & \vdots \\ y(T-1) & y(T-2) & \ldots & y(T-k) \\ y(T) & y(T-1) & \ldots & y(T-k+1) \\ 0 & y(T) & \ldots & y(T-k+2) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & y(T) \end{bmatrix} \quad (1.5)$$

are $(T+k) \times 1$ and $(T+k) \times k$ respectively. Thus the Yule-Walker estimators correspond to the ordinary least squares regression of $y(t)$ on $y(t-1),\ldots,y(t-k)$ for $t=1,\ldots,T+k$; with any unobserved y replaced by its expected value of zero.

Thus the Yule-Walker estimators suffer numerically by being the solution to normal equations for a linear model rather than being the result of applying a triangularization method (see Lawson and Hanson (1974), for example) to the design matrix X_k. This can be particularly important in time series because of ill-conditioning caused by the peculiar form of X_k. In addition, if the variance of y is large, then the effect of replacing the observations $y(1-k),\ldots,y(0)$ and $y(T+1),\ldots,y(T+k)$ that are on the ends of the observed data by zeros raises the possibility of obtaining poor estimators.

There have been three main attempts to alleviate the problems in Yule-Walker estimators. First, are what are often called least squares estimators (see Anderson (1971), p. 211). These are obtained by dropping the first and last k rows of $\underline{y}_k$ and X_k and then performing ordinary least squares regression.

These estimators seem to reduce end effects but cannot be obtained recursively for k=1,...,M and also are not guaranteed to lead to stable filters. The second method is the Burg (1968) algorithm which seems to moderate end effects, is recursive in nature, and leads to stable filters. However, it appears to be somewhat ad hoc in nature and its numerical properties are difficult to describe. The third method, due to Pagano (1972), consists of solving the normal equations (1.4) via the numerically stable modified Cholesky triangularization (Wilkinson (1967)) of $\hat{\Gamma}_M$.

The purpose of this paper is to show how the factors needed to find the modified Gram-Schmidt triangularization of X_M corresponding to Pagano's triangularization of $\hat{\Gamma}_M$ can be found recursively while only needing to store two $(T+M) \times 1$ vectors (rather than storing the entire X_M matrix). From this we find a design-matrix oriented algorithm, called the Toeplitz Gram-Schmidt algorithm, for finding the Yule-Walker estimators. We show that the Toeplitz Gram-Schmidt algorithm is roughly as fast as using Levinson's algorithm, while providing a unified framework for deriving other stable, more robust estimators. The algorithm allows calculation of diagnostics useful in measuring end effects as well as determining if particular observed values of y are exerting undue influence on the estimation procedure.

The algorithm is given in section 2, its numerical properties described in section 3, and its application for robust autoregression and autoregression diagnostics discussed in section 4.

2. THE TOEPLITZ GRAM-SCHMIDT YULE-WALKER ALGORITHM

We seek ordinary least squares estimators of $\underset{\sim}{\alpha}_k$, $k=1,\ldots,M$, for the linear models $\underset{\sim}{y}_k = -X_k\underset{\sim}{\alpha}_k + \underset{\sim}{\epsilon}$ without using the normal equations.

By Levinson's algorithm we can obtain all the $\hat{\alpha}_k(j)$ and $\hat{\sigma}_k^2$, from only $\hat{R}(0)$ and $\hat{\alpha}_k(k)$, $k=1,\ldots,M$, by

$$\hat{\sigma}_1^2 = \hat{R}(0) \ [1-\hat{\alpha}_1^2(1)] \tag{2.1}$$

$$\hat{\alpha}_k(j) = \hat{\alpha}_{k-1}(j) + \hat{\alpha}_k(k)\hat{\alpha}_{k-1}(k-j) \quad , \quad j=1,\ldots,k-1, \tag{2.2}$$

$$\hat{\sigma}_k^2 = \hat{\sigma}_{k-1}^2 \ [1-\hat{\alpha}_k^2(k)], \tag{2.3}$$

$k=2,\ldots,M$. Further if one only wants to estimate $\sigma_k^2/R(0)$, as is often the case, then only $\hat{\alpha}_k(k)$, $k=1,\ldots,M$ are required.

Thus the algorithm we propose consists of determining recursively $\hat{\alpha}_k(k)$, $k=1,\ldots,M$, by calculating at the kth step the kth and $(M+1)$st columns of the kth stage of the modified Gram-Schmidt decomposition (Björck (1967)), MGSD, of the $(T+M) \times (M+1)$ matrix $[X_M \vdots \underset{\sim}{y}_M]$. Then these $\hat{\alpha}_k(k)$'s are used in (2.1)-(2.3) to find all the desired estimators. Theorem 2.1 shows how the $\hat{\alpha}_k(k)$'s are calculated while Theorem 3.1 shows that the algorithm is indeed an application of the MGSD.

Define the lag operator L on the $(n \times 1)$ vector $\underset{\sim}{a} = (a(1),\ldots,a(n))^T$ to be $L(\underset{\sim}{a}) = (0, a(1),\ldots,a(n-1))^T$. Then we note that the columns of X_M are $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_M$ where $\underset{\sim}{x}_1 = L(\underset{\sim}{y}_M)$ and $\underset{\sim}{x}_j = L(\underset{\sim}{x}_{j-1})$, $j=2,\ldots,M$.

Theorem 2.1

Let $\underline{y}_{(1)} = \underline{y}_M$, $\underline{x}_{(1)} = L(\underline{y}_M) = \underline{x}_1$, and for $k = 1,\ldots,M$:

$$\hat{a}(k) = -\frac{\underline{y}_{(k)}^T \, \underline{x}_{(k)}}{\underline{x}_{(k)}^T \, \underline{x}_{(k)}} \quad ,$$

$$\underline{y}_{(k+1)} = \underline{y}_{(k)} + \hat{a}(k) \, \underline{x}_{(k)} \quad ,$$

$$\underline{x}_{(k+1)} = L(\underline{x}_{(k)} + \hat{a}(k)\underline{y}_{(k)}) \quad .$$

Then

a) $\hat{a}(k) = \hat{\alpha}_k(k)$, $k=1,\ldots,M.$

b) $\underline{y}_{(k)} = \underline{y}_M + \sum\limits_{j=1}^{k-1} \hat{\alpha}_{k-1}(j) \, \underline{x}_j$, $k=1,\ldots,M+1,$

c) $\underline{x}_{(k)} = \underline{x}_k + \sum\limits_{j=1}^{k-1} \hat{\alpha}_{k-1}(j) \, \underline{x}_{k-j},$ $k=1,\ldots,M+1,$

where any summation is zero if its upper limit is less than its lower limit.

Proof

We proceed by induction on k. For k=1, (b) and (c) are true by definition, while $\hat{a}(1) = -T\hat{R}(1)/T\hat{R}(0) = \hat{\alpha}_1(1)$. For k>1, we have by the induction hypothesis,

$$\underline{x}_{(k)} = L[\underline{x}_{(k-1)} + \hat{a}(k-1) \, \underline{y}_{(k-1)}]$$

$$= L[x_{(k-1)} + \sum_{j=1}^{k-2} \hat{\alpha}_{k-2}(j) x_{k-1-j}] + \hat{\alpha}_{k-1}(k-1) \; L[y_M + \sum_{j=1}^{k-2} \hat{\alpha}_{k-2}(j) x_j]$$

$$= x_k + \sum_{j=1}^{k-2} \hat{\alpha}_{k-2}(j) x_{k-j} + \hat{\alpha}_{k-1}(k-1) \; [x_1 + \sum_{j=1}^{k-2} \hat{\alpha}_{k-2}(j) x_j]$$

$$= x_k + \hat{\alpha}_{k-1}(k-1) x_1 + \sum_{j=1}^{k-2} [\hat{\alpha}_{k-2}(j) + \hat{\alpha}_{k-1}(k-1)\hat{\alpha}_{k-2}(k-1-j)] \; x_{k-j}$$

$$= x_k + \sum_{j=1}^{k-1} \hat{\alpha}_{k-1}(j) x_{k-j}$$

by (1.3). Similarly,

$$Y(k) = [y_M + \sum_{j=1}^{k-2} \hat{\alpha}_{k-2}(j) x_j] + \hat{\alpha}_{k-1}(k-1)[x_{k-1} + \sum_{j=1}^{k-2} \hat{\alpha}_{k-2}(j) x_{k-1-j}]$$

$$= y_M + \hat{\alpha}_{k-1}(k-1) x_{k-1} + \sum_{j=1}^{k-2} [\hat{\alpha}_{k-2}(j) + \hat{\alpha}_{k-1}(k-1)\hat{\alpha}_{k-2}(k-1-j)] x_j$$

$$= y_M + \sum_{j=1}^{k-1} \hat{\alpha}_{k-1}(j) \; x_j \qquad .$$

We have thus shown (a)-(c) being true for k-1 implies (b) and (c) are true for k and it remains to prove (a). To do so we need several well-known facts about the modified Cholesky decomposition of $\hat{\Gamma}_M$ and the Gram-Schmidt decomposition of X_M.

Let $X_M = Q_M R_M$ be the Gram-Schmidt decomposition of X_M, i.e R_M is an (M x M) unit upper triangular matrix and Q_M is a (T + M) x M orthogonal matrix having columns $q_1, \ldots, q_M$, i.e $Q_M^T Q_M = D_M$, a diagonal matrix. Let $\hat{\Gamma}_M = \hat{L}_M \hat{D}_M \hat{L}_M^T$ be the modified Cholesky decomposition of $\hat{\Gamma}_M$, i.e $\hat{L}_M$ is an

(M x M) unit lower triangular matrix and $\hat{D}_M$ is am (M x M) diagonal matrix.
Then $\hat{\Gamma}_M = \frac{1}{T} X_M^T X_M = \hat{L}_M \hat{D}_M \hat{L}_M^T = \frac{1}{T} R_M^T D_M R_M$, which by the uniqueness of the modified
Cholesky decomposition (Wilkinson (1967)), gives $\hat{L}_M = R_M^T$ and $\hat{D}_M = \frac{1}{T} D_M$.
Further (Pagano (1972)), with $\hat{\sigma}_0^2 = \hat{R}(0)$, we have

$$\hat{L}_M^{-1} = \begin{bmatrix} 1 & & & & \\ \hat{\alpha}_1(1) & 1 & & & \\ \hat{\alpha}_2(2) & \hat{\alpha}_2(1) & 1 & & \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot \\ \cdot & & & & & \cdot \\ \hat{\alpha}_{M-1}(M-1) & \hat{\alpha}_{M-1}(M-2) \ldots \hat{\alpha}_{M-1}(1) & & 1 \end{bmatrix} \quad , \tag{2.4}$$

$$\hat{D}_M = \text{Diag} \left(\hat{\alpha}_0^2, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_{M-1}^2 \right) \quad ,$$

and, defining $\hat{\underline{\theta}}_M = (\hat{\alpha}_1(1), \ldots, \hat{\alpha}_M(M))^T$,

$$\hat{\underline{\theta}}_M = -(Q_M^T Q_M)^{-1} Q_M^T \underline{y}_M \quad ,$$

i.e

$$\hat{\alpha}_k(k) = -\frac{\underline{q}_k^T \underline{y}_M}{\underline{q}_k^T \underline{q}_k} \quad , \qquad k=1,\ldots,M. \tag{2.5}$$

Since $X_M = Q_M R_M$ we have $Q_M^T X_M = Q_M^T Q_M R_M = D_M R_M$, an upper triangular matrix,
which gives

$$\underline{q}_k^T \underline{x}_j = 0 \quad , \qquad j<k.$$

Finally, $Q_M = X_M R_M^{-1} = X_M (C_M^{-1})^T$, _i.e_ by (2.4),

$$q_k = \underset{\sim}{x}_k + \sum_{j=1}^{k-1} \hat{\alpha}_{k-1}(j)\, \underset{\sim}{x}_{k-j} \quad ,$$

which we have shown under the induction hypothesis is $\underset{\sim}{x}_{(k)}$.

Thus,

$$\hat{a}(k) = -\frac{\underset{\sim}{x}_{(k)}^T \underset{\sim}{y}_{(k)}}{\underset{\sim}{x}_{(k)}^T \underset{\sim}{x}_{(k)}} = -\frac{q_k^T[\underset{\sim}{y}_M + \sum_{j=1}^{k-1} \hat{\alpha}_{k-1}(j)\underset{\sim}{x}_j]}{q_k^T q_k}$$

$$= -\frac{q_k^T \underset{\sim}{y}_M}{q_k^T q_k} = \hat{\alpha}_k(k)$$

by (b) and (2.5).

<div align="right">QED.</div>

Corollary 2.1

The vectors $\underset{\sim}{y}_{(k)}$ and $\underset{\sim}{x}_{(k)}$ are the forward and backward prediction errors one would obtain using a memory (k-1) linear prediction filter assuming $y(1-k) =\ldots=y(0)=y(T+1)=\ldots=y(T+k)=0$, _i.e_ if we denote by $y_{(k)}^j$ and $x_{(k)}^j$ the jth elements of $\underset{\sim}{y}_{(k)}$ and $\underset{\sim}{x}_{(k)}$, then for $j=1,\ldots,T$,

$$y_{(k)}^j = -\sum_{\ell=1}^{k-1} \hat{\alpha}_{k-1}(\ell) y(j-\ell)$$

$$x_{(k)}^j = -\sum_{\ell=1}^{k-1} \hat{\alpha}_{k-1}(\ell) y(j+\ell)$$

where $\hat{\underset{\sim}{\alpha}}_{k-1}$ is the value of $\underset{\sim}{a} = (a(1),\ldots,a(k-1))^T$ minimizing both

$$S_F(\underline{a}) = \sum_{j=1}^{T} \{y(j) + \sum_{\ell=1}^{k-1} a(\ell)y(j-\ell)\}^2$$

and

$$S_B(\underline{a}) = \sum_{j=1}^{T} \{y(j) + \sum_{\ell=1}^{k-1} a(\ell)y(j+\ell)\}^2 \quad ,$$

assuming $y(1-k)=\ldots=y(0)=y(T+1)=\ldots=y(T+k)=0$. Further,

$$\underline{x}^T(k)\underline{x}(k) = S_B(\hat{\underline{\alpha}}_{k-1}) = S_F(\hat{\underline{\alpha}}_{k-1}) = \underline{y}^T(k)\underline{y}(k) \quad , \tag{2.6}$$

$k=1,\ldots,M+1$.

The Toeplitz Gram-Schmidt Yule-Walker algorithm for finding $\hat{\alpha}_k(j),\hat{\sigma}_k^L$, $1\leq j\leq k\leq M$, consists then of the following steps:

i) $\quad \underline{y}(1) = \underline{y}_M, \underline{x}(1)=L(\underline{y}_M), \hat{\sigma}_o^2 = \hat{R}(0)$ $\tag{2.7}$

ii) $\quad$ for $k=1,\ldots,M$: $\tag{2.8}$

$$\hat{\alpha}_k(k) = - \frac{\underline{y}^T(k)\underline{x}(k)}{\underline{x}^T(k)\underline{x}(k)}$$

$$\hat{\sigma}_k^2 = \hat{\sigma}_{k-1}^2 \ (1-\hat{\alpha}_k^2(k))$$

$$\underline{y}(k+1) = \underline{y}(k) + \hat{\alpha}_k(k)\underline{x}(k)$$

$$\underline{x}(k+1) = L(\underline{x}(k) + \hat{\alpha}_k(k)\underline{y}(k))$$

iii) $\quad$ use (2.1), (2.2) to obtain the other $\hat{\alpha}_k(j)$'s.

3. NUMERICAL PROPERTIES OF THE ALGORITHM

3.1 Arithmetic Operations and Required Storage

To obtain $\hat{\theta}_M$ there are $2(M-1)$ updates of the form $\underline{a} - c\underline{b}$ for $(T+M) \times 1$ vectors $\underline{a}$ and $\underline{b}$ and $2M$ inner products of $(T+M) \times 1$ vectors. Counting a multiplication and addition as an operation, there are thus $(T+M)(4M-2)$ operations to obtain $\hat{\theta}_M$.

To use Levinson's algorithm, $\hat{R}(0),\ldots,\hat{R}(M)$ must first be calculated, which takes approximately TM operations if inner products are used or $2T \log_2 T$ operations if the Fast Fourier Transform is used.

In terms of speed then the algorithm of Theorem 2.1 is slightly slower than using Levinson's algorithm. Neither of the algorithms requires significant storage; approximately $2(T+M)$ locations for the Toeplitz Gram-Schmidt algorithm versus $T+M$ for Levinson's algorithm when the data vector is stored.

3.2 Error Analysis

To determine the numerical accuracy of the proposed algorithm we shall first prove that it is a method of finding the matrix Q and the last column of R in the MGSD of $[X_M \vdots \underline{y}_M]$ using only two vectors of length $T+M$.

Let A be a full rank $(n \times p)$ matrix having columns $\underline{a}_1,\ldots,\underline{a}_p$. The MGSD algorithm is a numerically stable (Björck (1967)) method of determining the Gram-Schmidt factors $Q=(\underline{q}_1,\ldots,\underline{q}_p)$ and R of A in p stages. At the kth stage, the kth column of Q and the kth row of R are determined. Conceptually, one defines the sequence of matrices $Q^{(0)} = A$, $Q^{(1)},\ldots,Q^{(p)} = Q$, with $Q^{(k)}$ having columns $\underline{q}_1^{(k)},\ldots,\underline{q}_p^{(k)}$, by:

$$q_j^{(k)} = \begin{cases} q_j^{(k-1)} & , \quad j \leq k \\ \\ q_j^{(k-1)} + R_{kj} \, q_k^{(k)} & , \quad j > k \end{cases}$$

where R_{kj} is the ratio of inner products

$$R_{kj} = - \frac{(q_j^{(k-1)})^T \, q_k^{(k)}}{(q_k^{(k)})^T \, q_k^{(k)}}$$

Then $q_k = q_k^{(k)}$ and one can actually accumulate $Q^{(0)}$, $Q^{(1)}$,...,$Q^{(p)}$ in a single matrix.

We have

Theorem 3.1

The vectors $x_{(k)}$ and $y_{(k)}$ in Theorem 2.1 are the kth and $(M+1)st$ columns of $Q^{(k)}$ in the MGSD of matrix $[X_M \colon y_M]$.

Proof

Parts (b) and (c) of Theorem 2.1 show that for $k>1$, $y_{(k)}$ and $x_{(k)}$ and the ordinary least squares residuals of the regression of y_M on $-x_1,\ldots,-x_{k-1}$ and x_k on $-x_{k-1},\ldots, -x_1$ respectively and thus (see Clayton (1971)) they are $q_{M+1}^{(k)}$ and $q_k^{(k)}$ respectively. For $k=1$,

$$R_{12} = - \frac{x_2^T \, x_1}{x_1^T \, x_1} = - \frac{y_M^T \, x_1}{x_1^T \, x_1} = R_{1,M+1} = \hat{a}_1(1).$$

QED

Since the Toeplitz Gram-Schmidt algorithm is a subset of the MGSD we can use the results of Björck (1967). We use the techniques and notation introduced by Wilkinson (1965).

Using equation (5.8) from Björck (1967) and assuming that all inner products are accumulated in double precision, we have

$$|| fl(\underline{y}_{(M)}) - \underline{y}_{(M)} || \leq 1.9 \, M^{\frac{1}{2}} \, (M+1) 2^{-t} \sqrt{R(0)} \quad , \tag{3.1}$$

where 2^{-t} is the machine precision of the computer being used to perform the calculations. This proves that the Toeplitz Gram-Schmidt algorithm gives a numerically stable prewhitening filter. By symmetry, a formula identical to (3.1) would apply for $\underline{x}_{(M)}$.

The roundoff error for $\hat{\alpha}_k(k)$ can be obtained by letting $\tilde{\alpha}_k(k)$ be the exact solution (no rounding errors) using the inexact $fl(\underline{x}_{(k)})$ and $fl(\underline{y}_{(k)})$ so that

$$| fl(\hat{\alpha}_k(k)) - \hat{\alpha}_k(k) | \leq | fl(\hat{\alpha}_k(k)) - \tilde{\alpha}_k(k) |$$

$$+ | \tilde{\alpha}_k(k) - \hat{\alpha}_k(k) | \quad .$$

Using a perturbation bound on $\tilde{\alpha}_k(k)$ and equation (4.11) in Björck, one obtains

$$| fl(\hat{\alpha}_k(k)) | - \tilde{\alpha}_k(k) | < (2.01 | \hat{\alpha}_k(k) | + .01 \, \hat{\sigma}_k^2) \, 2^{-t},$$

$$| \tilde{\alpha}_k(k) - \hat{\alpha}_k(k) | < 8k^{\frac{1}{2}} \, (k+1) \sqrt{R(0)} \quad 2^{-t} \, ,$$

and hence

$$|f1(\hat{\alpha}_k(k)) - \hat{\alpha}_k(k)| < (2.01|\hat{\alpha}_k(k)| + .01 \hat{\sigma}_k^2 + 8k^{\frac{1}{2}}(k+1) \sqrt{R(0)})2^{-t} .$$

From this it is concluded that the algorithm is stable.

To illustrate the numerical accuracy of the algorithm, we use the example of a constant $y(t)$ (see Davis and Qualls (1977)). For $T=1000$ and $M=20$, $\hat{\Gamma}_M$ has condition number $K(\hat{\Gamma}_M) = 505$. The results of using Levinson's algorithm and the Toeplitz Gram-Schmidt algorithm (multiplied by 2^t to make them reasonably machine independent) are presented in Table 3.1.

Table 3.1

	$||f1(\hat{\underline{\alpha}}_M) - \hat{\underline{\alpha}}_M|| 2^t$	K
Levinson	1168	505
Toeplitz Gram-Schmidt	.41	22.5

4. AUTOREGRESSION DIAGNOSTICS AND ROBUST AUTOREGRESSION

In this section we consider the application of methods in ordinary regression analysis for robustness and detecting influential observations in the autoregression situation. We will show that the Toeplitz Gram-Schmidt algorithm provides a unified framework for carrying out this analysis.

4.1 Detecting Influential Observations

In the usual full rank, n-observation, p-parameter, linear model $\underset{\sim}{y} = X\beta + \underset{\sim}{\varepsilon}$, the diagonal elements of the hat matrix $H = X(X^TX)^{-1}X^T$ are often used (Hoaglin and Welsch (1978)) to determine if undue weight is given any particular observation in finding the vector of predicted values $\hat{\underset{\sim}{y}} = X\hat{\beta} = X(X^TX)^{-1}X^T\underset{\sim}{y} = H\underset{\sim}{y}$.

If $X = QR$ where $Q^TQ = D = \text{Diag}(d_1,\ldots,d_p)$, we have $H = QR[R^{-1}D^{-1}R^{-T}]R^TQ^T = QD^{-1}Q^T$. Thus

$$H_{jj} = \sum_{\ell=1}^{p} Q_{j\ell}^2/d_\ell \qquad , j=1,\ldots,n \quad .$$

In the autoregression case then these hat diagonals can be easily accumulated in the Toeplitz Gram-Schmidt Algorithm. If we define $d_{(k)} = \underset{\sim}{x}_{(k)}^T \underset{\sim}{x}_{(k)}$, $k=1,\ldots,M+1$, and

$$h_{(1)}^j = (x_{(1)}^j)^2/d_{(1)} \qquad , j=1,\ldots,T+M,$$

$$h_{(k)}^j = h_{(k-1)}^j + (x_{(k)}^j)^2/d_{(k)} \quad , \quad j=1,\ldots,T+M, k=2,\ldots,M+1,$$

then $h_{(k)} = (h_{(k)}, \ldots, h_{(k)}^{T+M})^T$ are the hat diagonals of the regression of y_M on $-x_1, \ldots, -x_{k-1}$. As such they can be used both for detecting influential observed y's and for measuring end effects. Note that calculation of the $h_{(k)}$ can be done easily by just calculting $d_{(k)}$ and $h_{(1)}$ in (2.7) and $d_{(k)}, h_{(k)}$ in (2.8).

4.2 Alternative Estimators

A wide variety of possibly robust estimators of autoregressive parameters are suggested by the results of Theorem 2.1 and Corollary 2.1, i.e the Yule-Walker $\hat{\alpha}_k(k)$ is the regression coefficient of the regression through the origin of the forward prediction errors $y_{(k)}$ on the backward prediction errors $x_{(k)}$. Thus any of the M-estimators (Huber (1973) could be used rather than ordinary least squares to obtain estimators $\tilde{\alpha}_1(1), \ldots, \tilde{\alpha}_M(M)$. Then these would be used instead of $\hat{\alpha}_1(1), \ldots, \hat{\alpha}_M(M)$ in (2.1)-(2.3), thus giving what could be called ℓ_2-M autoregressive parameter estimators, ℓ_2 for the use of Levinson's algorithm and M for the use of M-estimators. We note that such M-estimators probably need not lead to $\tilde{\alpha}_k(k)$'s that are less than one in absolute value (A necessary and sufficient condition for stability of resulting prediction filters) but this may actually be a benefit in that it would indicate that the original time series is not stationary.

Now since $x_{(k)}^T \underline{1} = y_{(k)}^T \underline{1} = 0$ if $y_M^T \underline{1} = 0$ we have also that $\hat{\alpha}_k(k)$ is the negative of the ordinary sample correlation coefficient of the forward and backward prediction errors for order (k-1). Thus again one could use robust methods of estimating correlation coefficients (Huber (1977)) to obtain other estimators $\tilde{\alpha}_k(k)$.

4.3 Relationship of the Toeplitz Gram-Schmidt and Burg Algorithms

An important method of autoregressive parameter estimation is the Burg (1968) algorithm. With the framework provided by the Toeplitz Gram-Schmidt algorithm it is easy to describe what the Burg algorithm does. It uses the same updating equations but uses

$$\tilde{\alpha}_k(k) = - \frac{\sum\limits_{j=1}^{T-k} y_{(k)}^{j+k} \; x_{(k)}^{j+k}}{\frac{1}{2}\left[\sum\limits_{j=1}^{T-k} (x_{(k)}^{j+k})^2 + \sum\limits_{j=1}^{T-k} (y_{(k)}^{j+k})^2 \right]}$$

instead of $\hat{\alpha}_k(k)$. By (2.6) note that $\tilde{\alpha}_k(k)$ is identical to $\hat{\alpha}_k(k)$ except that the summations range from 1 to T-k instead of from 1 to T+M. Note also that $\tilde{\alpha}_k(k)$ is the value of a minimizing

$$\sum\limits_{j=1}^{T-k} \{(y_{(k)}^{j+k} + a \; x_{(k)}^{j+k})^2 + (x_{(k)}^{j+k} + a \; y_{(k)}^{j+k})^2\} \; ,$$

i.e the sum of squares of error in regressing the "observed" forward prediction errors on the "observed" backward prediction errors and vice versa, i.e the Burg algorithm never uses a y outside the range of observed values. We note further that Stuart (1955) has shown that if $(x_1,y_1),\ldots,(x_n,y_n)$ is a random sample from a bivariate normal distribution with zero means and equal but unknown variances, then

$$r' = \frac{\sum\limits_{i=1}^{n} x_i y_i}{\frac{1}{2} [\sum\limits_{i=1}^{n} x_i^2 + \sum\limits_{i=1}^{n} y_i^2]}$$

is an asymptotically efficient estimator of the correlation coefficient of x and y. Thus the Burg estimator of $\alpha_k(k)$ can also be interpreted as the negative of a correlation coefficient estimator.

Thus if k is large relative to T and R(0) is small the Burg algorithm may be inferior to Yule-Walker estimation, while if R(0) is large it will probably be superior because of end effects in Yule-Walker estimation. Finally we note that the $\tilde{\alpha}_k(k)$ are guaranteed to be less than one in absolute value.

REFERENCES

Akaike, H. (1969), "Fitting Autoregressive Models For Prediction," _Annals_
of the Institute of Statistical Mathematics, 21, 243-247.

Anderson, T. W. (1971), _The Statistical Analysis of Time Series_, New York:
John Wiley.

Björck, A. (1967), "Solving Linear Least Squares Problems by Gram-Schmidt
Orthogonalization," _BIT_, 7, 1-21.

Burg, J. P. (1968), "A New Analysis Technique for Time Series Data,"
presented at NATO Advanced Study Conference on Signal Processing,
Enschede, Netherlands.

Clayton, D. G. (1971), _"Gram-Schmidt Orthogonalization of Vectors,"_
Applied Statistics, 20, 335-338.

Davis, H. T. and Qualls, C. (1977), "Time Series Software Comparisons,"
1977 Proceedings of the Statistical Computing Section, ASA,
33-37.

Hoaglin, D. C. and Welsch, R. E. (1978), "The Hat Matrix in Regression and
ANOVA," _American Statistician_, 32, 17-22.

Huber, P. J. (1973), "Robust Regression: Asymptotics, Conjectures, and
Monte Carlo," _Annals of Statistics_, 1, 799-821.

_____. (1977), _Robust Statistical Properties_, Philadelphia; SIAM.

Lawson, C. L. and Hanson, R. J. (1974), _Solving Least Squares Problems_, Englewood Cliffs, New Jersey: Prentice-Hall.

Levinson, N. (1947), "The Wiener RMS Error Criterion in Filter Design and Prediction," _Journal of Mathematical Physics_, 25, 261-278.

Pagano, M. (1972), "An Algorithm for Fitting Autoregressive Schemes," _Applied Statistics_, 21, 274-281.

Parzen, E. (1961), "An Approach to Time Series Analysis," _Annals of Mathematical Statistics_, 32, 951-989.

_____. (1974), "Some Recent Advances in Time Series Modeling," _IEEE Transactions on Automatic Control_, AC-19, 723-730.

Stuart, A. (1955), "A Paradox in Statistical Estimation," _Biometrika_, 42, 527-529.

Ulrych, T. L. and Bishop, T. N. (1975), "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," _Review of Geophysics and Space Physics_, 14, 183-200.

Wilkinson, J. H. (1965), _The Algebraic Eigenvalue Problem_, Oxford: Oxford Press.

_____. (1967), "The Solution of Ill-Conditioned Linear Equations," in _Mathematical Methods for Digital Comptuers_, eds. H. Ralston and H. S. Wilf, New York: John Wiley.